Normalization strategies in multi-center radiomics abdominal MRI: systematic review and meta-analyses

IEEE Open Journal of Engineering in Medicine and Biology

Emerging Topics

Normalization strategies in multi-center radiomics abdominal MRI: systematic review and meta-analyses

Panic J.*, *Member, IEEE*, Defeudis A., Balestra G., *Member, IEEE*, Giannini V. (†), *Member, IEEE*, Rosati S. (†), *Member, IEEE*

***Abstract*— *Goal:* Artificial intelligence applied to medical image analysis has been extensively used to develop non-invasive diagnostic and prognostic signatures. However, these imaging biomarkers should be largely validated on multi-center datasets to prove their robustness before they can be introduced into clinical practice. The main challenge is represented by the great and unavoidable image variability which is usually addressed using different pre-processing techniques including spatial, intensity and feature normalization. The purpose of this study is to systematically summarize normalization methods and to evaluate their correlation with the radiomics model performances through meta-analyses. This review is carried out according to the PRISMA statement: 4777 papers were collected, but only 74 were included. Two meta-analyses were carried out according to two clinical aims: characterization and prediction of response. Findings of this review demonstrated that there are some commonly used normalization approaches, but not a commonly agreed pipeline that can allow to improve performance and to bridge the gap between bench and bedside.**

***Index Terms*—abdominal MRI, artificial intelligence, multi-center database, normalization, radiomics.**

***Impact Statement*— Our research demonstrated the lack of a standardized abdominal pre-processing pipeline to normalize MRI images and features across centers. As a consequence, we proved the need to select the most suitable normalization methods depending on image characteristics and clinical questions.**

I. Introduction

CURRENTLY, the role of medical imaging is evolving from being mainly a diagnostic tool to gaining a central role in the context of personalized precision medicine [1]. This paradigmatic shift was made possible by the development of radiomics, which allows the extraction, from medical images, of quantitative features providing useful information on diagnosis and prognosis [1], [2]. Despite the encouraging results provided by recent studies [3]–[6], there is still no radiomics-based systems used in clinical practice for abdominal imaging. This is mainly due to the lack of multi-center clinical trials for both system development and validation [7]. The acquisition of images from several institutions is surely complex for many technical reasons, in addition to legal, ethical and administrative issues [8]. From the technical point of view, the most relevant obstacle is represented by the unavoidable high image intensity distribution variability due to different scanners, acquisition protocols, reconstruction settings and the patients' characteristics. This issue is particularly relevant when a multi-center dataset is used, and multi-center validation is compelling for developing robust, reproducible, and statistically relevant results.

Different standardization guidelines have been proposed to address the above-mentioned problem in the case of computer tomography (CT) and positron emission tomography (PET)/CT imaging, while they are not available for Magnetic Resonance Imaging (MRI) [9]. However, signal intensities of MRI are non-standardized and highly dependent on manufacturer and acquisition protocol parameters [10]. Therefore, bigger efforts are needed to solve problems related to low repeatability and reproducibility.

Currently, different pre-processing algorithms have been implemented to reduce the variability [11]–[13], especially in studies involving retrospective databases when it is not feasible to use standardized imaging acquisition protocols [1]. The most used approach is normalization, including a set of techniques in which values are shifted and/or rescaled, and that could be applied to parameters related to different image characteristics, i.e., physical dimensions (spatial normalization), pixels intensity (intensity normalization) and radiomics features (feature normalization).

The aim of this study is to systematically review normalization approaches applied to multi-center abdominal MRI, to assess whether it is possible to provide guidelines or evidences about the performances of the most frequently used methods. Even if this topic was widely addressed for the brain [14], [15], to the best of our knowledge, there is no systematic review regarding the abdominal area in the literature. As a secondary endpoint, we conducted different meta-analyses to

*P.J. is with University of Turin, Department of Surgical Science, and Polytechnic of Turin, Department of Electronics and Telecommunications, (address: Corso Duca degli Abruzzi, 24,10129 Turin, Italy; correspondence e-mail: jovana.panic@polito.it). D.A., G.V. are with University of Turin, Department of Surgical Science, Turin, Italy, and with Candiolo Cancer Institute, FPO-IRCCS, Candiolo (TO), Italy. S.R., and B.G. are with Polytechnic of Turin, Department of Electronics and Telecommunications, Turin, Italy. (†)G.V. and R.S. contributed equally to this work and share last authorship.

understand the impact of normalization methods on radiomics models, according to their aim.

II. Methods

A. Search strategy

This review was carried out according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement. Relevant articles were identified by searching three databases: PubMed, Web of Science and Scopus. The query used for the literature search was "MRI" AND "multicenter" AND ("database" OR "trial" OR "standardization" OR "normalization") AND ("radiomics" OR “Artificial Intelligence” OR “Machine Learning” OR "Deep Learning").

Literature searching, study identification, and data extraction from eligible studies were performed by one investigator with experience in abdominal radiomics field research (J.P.).

B. Eligibility criteria

Searched studies had to further come across the following eligibility criteria to be incorporated in the present review: (i) written in English; (ii) AI radiomics-based studies; (iii) published between January 2012 (when the radiomics definition has been first published by Lambin *et al.*[2]) and December 2022; (iv) based on a multi-center MRI database; (v) original research work published in a peer-reviewed journal.

C. Exclusion criteria

Studies were excluded based on any of the following criteria: (i) papers assessing image quality, (ii) papers describing only harmonization or standardization acquisition pipelines without using them for the development of a radiomics signature or AI model for detection/characterization/prognosis, (iii) papers describing challenges or online databases, and (iv) not clearly specifying the pre-processing step.

D. Papers analysis

For all included papers, the following information was collected (when available either on the manuscript or in supplementary materials): organ of interest, clinical aim, publication year, database details, normalization approach and method, radiomics features extracted, developed radiomics model, distinguishing between Machine and Deep Learning (ML and DL), and performances on the validation set. We grouped clinical aims in three categories: (i) detection as the capability to detect the pathologic tissue, i.e., its presence, and, if confirmed, its localization on the medical image; (ii) characterization as the ability to predict the clinical outcome; and (iii) prediction of therapy response, only appliable on pathologic tissues.

E. Meta-Analysis

We conducted different meta-analyses on the model performances, according to the aim, i.e., detection, characterization, response to therapy. We considered the area under the curve (AUC) and its Confidence Interval (CI) as the performance metric, which was the most frequently evaluated parameter among the studies. We excluded papers that did not report AUC and neither its standard deviation, standard error or CI values, and that weren’t validated on an external dataset, since this does not determine model reproducibility and generalizability to new and different patients [16].

The random-effects model was used to calculate the pooled AUC and to produce the forest plot. Cochran’s Q and the I^2 statistic were calculated. Cochran's Q statistic tests the studies' heterogeneity, under the null hypothesis H_0 that all studies are homogeneous (a *p-value*<0.05 was considered statistically significant). The I^2 value was used to quantify heterogeneity, providing an estimate of the percentage of variability among included studies: values of 25% and less are usually considered to be low, 25% - 50% moderate and above 75% are considered high. In case of detected heterogeneity, a moderator analysis was carried out by dividing studies into subgroups according to the organ, for each aim. Subgroups containing only one paper were excluded from the meta-analysis. The weight of each study was calculated with the inverse variance method, in which the weight given to each study is inverse of the variance of the effect estimate, minimizing the uncertainty of the pooled effect estimate [17]. The meta-analysis and the described statistical analyses were performed using R and the *metafor* package [18].

III. Results and Discussions

A. Papers collection

During the identification step, 4777 papers were collected from three search databases. Among them 1574 were duplicate records, therefore they were excluded resulting in a total of 3203 papers. During the screening step, 3069 were excluded

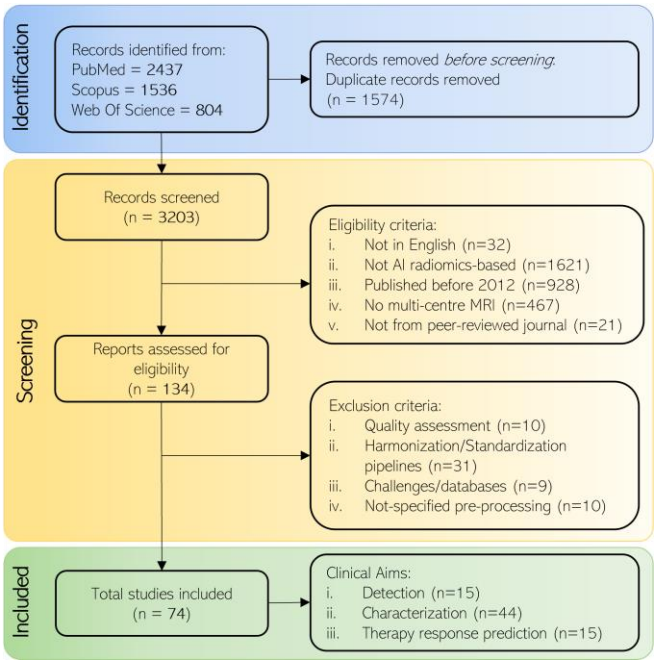

Fig. 1: Flow chart of the selection process of the studies included in the present review. At first, the author searched the 3 databases to identify the relevant articles for the study (Identification). Then, she screened the initially obtained studies considering the eligibility and exclusion criteria (Screening), and finally, selected the articles used for the study, based on the research objectives (Included).

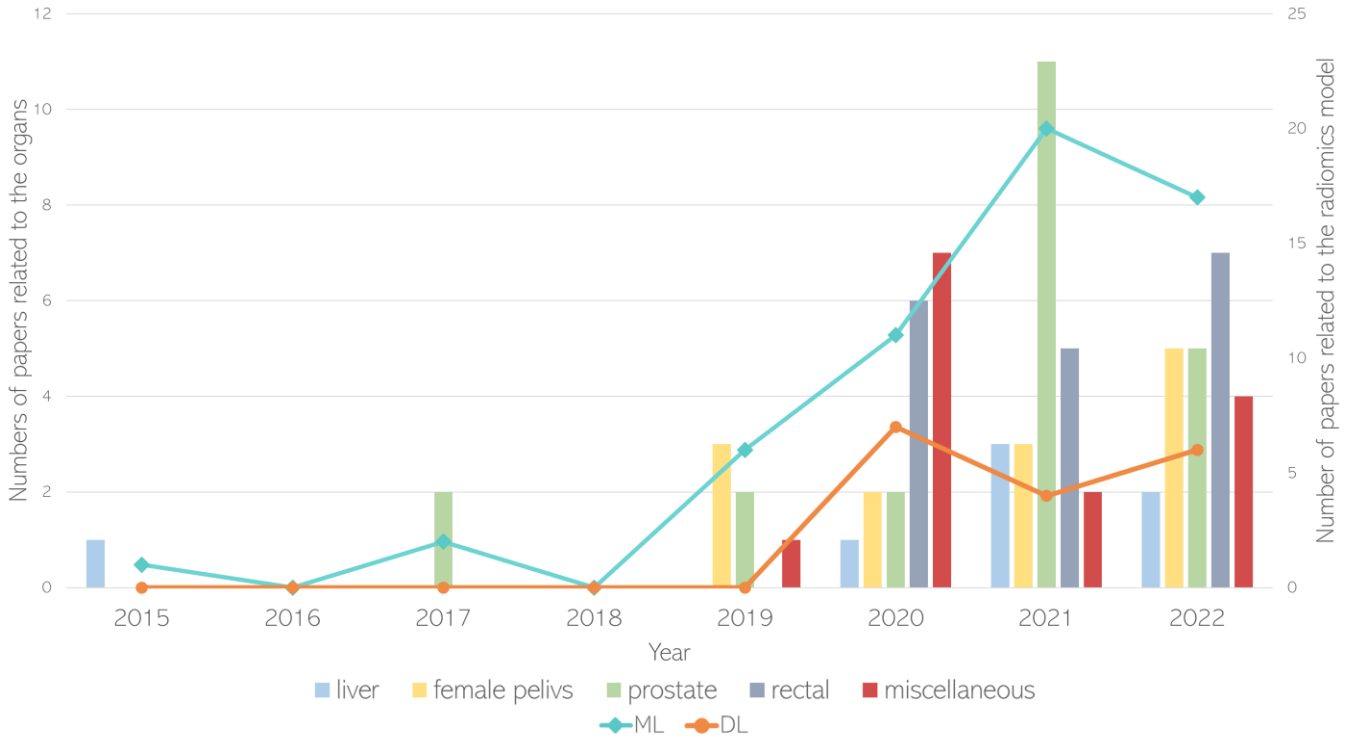

Fig. 2: This combined graph shows the distribution of the number of papers related to the abdominal organs (bars) and radiomics models (lines) from 2015 to 2022.

as they did not meet the eligibility criteria. Of the remaining 134 papers, 60 were eliminated after reviewing the study context and design. Finally, 74 papers were included in the final analysis. Fig. 1 reports the selection process and the number of papers per aim. All comparison tables reporting the normalization approaches, features, models, and performances are included in the supplementary materials, one for each organ.

B. *Clinical analysis*

All included papers have been published since 2015, with an increase from 2019. The most studied organs are the prostate [3], [19]–[40] (S-Table I), female pelvis [4], [41]–[52] (S-Table II), and rectum [5], [53]–[69] (S-Table III), while a lower number of publications were related to the liver [6], [70]–[75] (S-Table IV), and a miscellaneous group of organs including kidney [76]–[80], bladder [81], [82], pancreas [83], [84], and soft abdominal tissues [85]–[89] (S-Table V). As shown in fig. 2, the first study applying ML on multi-center databases was published in 2015 for the characterization of liver fibrosis, classifying them into five groups, ranging from no-fibrosis (0) to cirrhosis (5) [70]. Subsequently, two studies were published in 2017 developing ML models for prostate cancer detection: one assessing the differences between the transactional and peripherical zone of the prostate [40], while the other providing useful information for radiotherapy dose differentiation treatments [36].

The turning year, in which the number of papers showed a big rise also in other abdominal organs, was 2020. This might be due to the increasing number of public databases [90]–[92] and/or collaboration between institutions. Moreover, also a growing interest related to DL models was observed in the literature from this year. Considering the clinical question, as shown in the supplementary materials, ML is used to non-invasively characterize pathological tissues in 67% of the papers (38/57), predict the therapy response in 25% (14/57) and detect the disease in 8% (5/57), while DL is mostly used for detection (10/17 of papers) while only 6/17 (35%) and 1/17 (6%) of papers focus on characterization and therapy response, respectively.

C. *Technical analysis*

Fig. 3 shows that 77% of papers developed ML radiomics systems, while the remaining 23% were based on DL. In both cases, most of the studies (>70%) used the multi-center database to externally validate the model. Focusing on the ML systems, we analyzed the extracted radiomics feature groups. As shown by the bar diagram, the Gray-Level Co-Occurrence Matrix (GLCM) is the most used one, followed by the Gray-Level Run Length Matrix (GLRLM) and the First Order. Interestingly, since 2021 deep learning features [93] were introduced, e.g., Hiremath *et al.*[22] and Liu *et al.* [94].

Of note, since 2019, 15/74 papers include clinical features, available from routine practice, into their model in order to increase the predictive potential usefulness (S-Table I-V).

1) *Normalization approaches*

The following normalization approaches were used on MRI datasets:

- *Spatial normalization*: a process that changes the spatial characteristics of the image, e.g., the pixel's resolution, the Field of View (FOV), sequences orientation, etc. All methods apply geometrical transformations (e.g., resampling to obtain a fixed

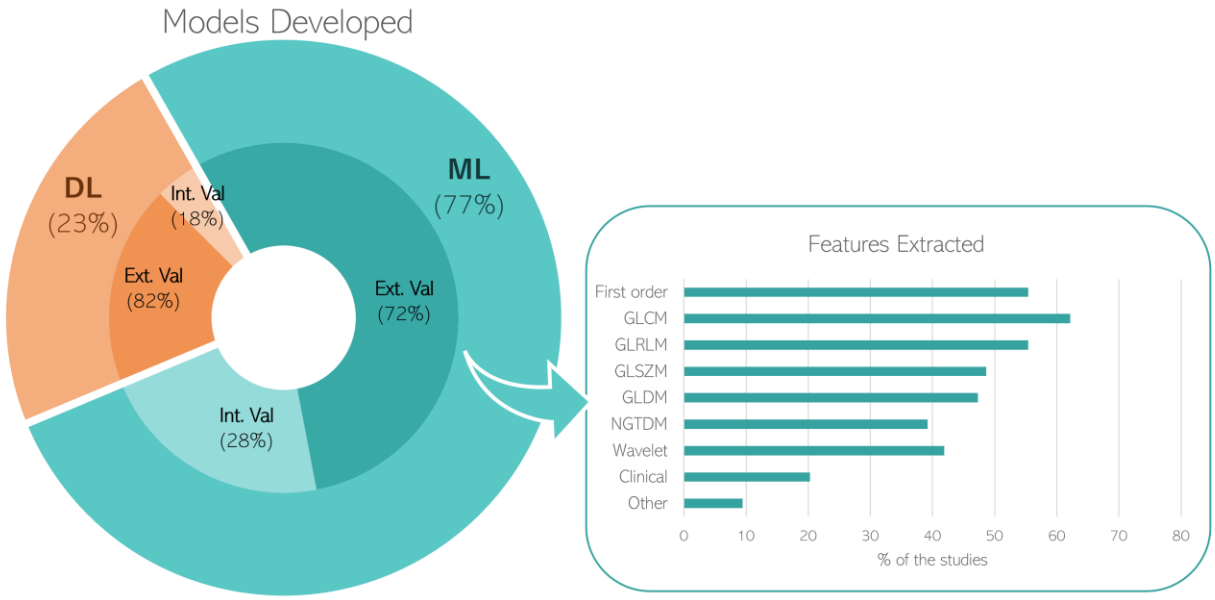

Fig. 3: The pie graph (on the left) shows the distributions of the models developed among the included papers. Specifically, the percentage related to the internal (Int.) and external (Ext.) validations are presented for both Machine Learning (ML) (blue) and Deep Learning (DL) (orange). The bar diagram (on the right) shows the percentage of different radiomics feature groups extracted in the ML studies.

pixel resolution), which do not need normalization parameters. This approach allows obtaining images including almost the same anatomical regions, thus avoiding histogram mismatches due to different FOVs rather than signal intensity dissimilarities. However, spatial modifications may alter the organs' morphology, undermining the clinical significance.

- *Intensity normalization*: a process that rescales the pixel intensity values to the same range. Most methods allow evaluating the normalization parameters independently on the sequences, except those which need a normalization reference, e.g., *hist_norm* [95], and *NyulUdupa* [96]. At the same time, modifying the histogram distribution may alter the clinical significance of both healthy and pathological tissues.
- *Feature normalization*: a process that rescales values of the radiomics feature extracted to the same range. On one hand, all methods allow reducing the differences between the features, without altering the clinical significance. On the other hand, a training set is needed to evaluate the normalization parameters.

Among all, intensity normalization is applied on 61/74 (82%) papers, being the most frequent, while the feature and spatial normalizations on 27/74 (36%) and 48/74 (65%) papers, respectively. In 52/74 studies the images underwent more than one normalization approach: 6/74 (8%) both intensity and feature, 34/74 (46%) spatial and intensity, 2/74 (3%) spatial and feature, and 10/74 (14%) all three approaches. Colored boxes in fig. 4 list the normalization methods used for each approach.

In general, three methods are currently used for spatial normalization: *registration* to align the sequences and reduce the motion artefacts, *resampling* to obtain the same resolution, and *resizing* to obtain the same size for all images. The most commonly applied method is *resampling* (31/48), which has been used individually in 12/16 studies on the prostate, 4/5 on the liver, and 6/9 in the miscellaneous group. Moreover, it has been applied in combination with at least another spatial normalization method in 14/48 papers, i.e., *resampling* and *resizing* on the female pelvis (4/9), and the rectum (3/9). *Registration* is the least frequently used method (10/48) among all organs. For intensity normalization, the two most used methods are the standardization (*z-score*) of the intensity distribution, which is applied on 21/61, and *custom* algorithms on 31/61, including normalization using pre-defined values, discretization of the distributions, the use of Advanced Normalization Tools (ANTs) or filters, the application of DL methods for harmonizing the image, or the combination of more than one method. The remaining 9/61 papers applied one of the following methods: *min-max* or *3sigma* scaling and mean centering (*mean-cent*), and histogram normalization (*hist-norm*) which matches the original distribution to a histogram reference (e.g., healthy subject, other organs, etc.). Most papers related to the prostate (10/20) and miscellaneous (8/10) have developed custom methods for the intensity normalization, while the *z-score* is frequently applied to female pelvis (5/10) and rectal (7/16), and *hist-norm* on the liver (2/5). More details are in S-Table I-V. Regarding feature normalization, only three methods have been found in the included papers and are almost equally used: i.e., rescaling according to the minimum and maximum feature values (*min-max*) is applied on 6/27 papers, standardization (*z-score*) on 9/27, and harmonization using *ComBat* [12] on 5/27. The first two methods are very easy to implement and could be applied on different centers without a training phase, i.e., using the same normalization parameters derived from the previously included dataset. Conversely, the *ComBat* method can only normalize features across different centers by using values computed on a subset of data from each center that will be

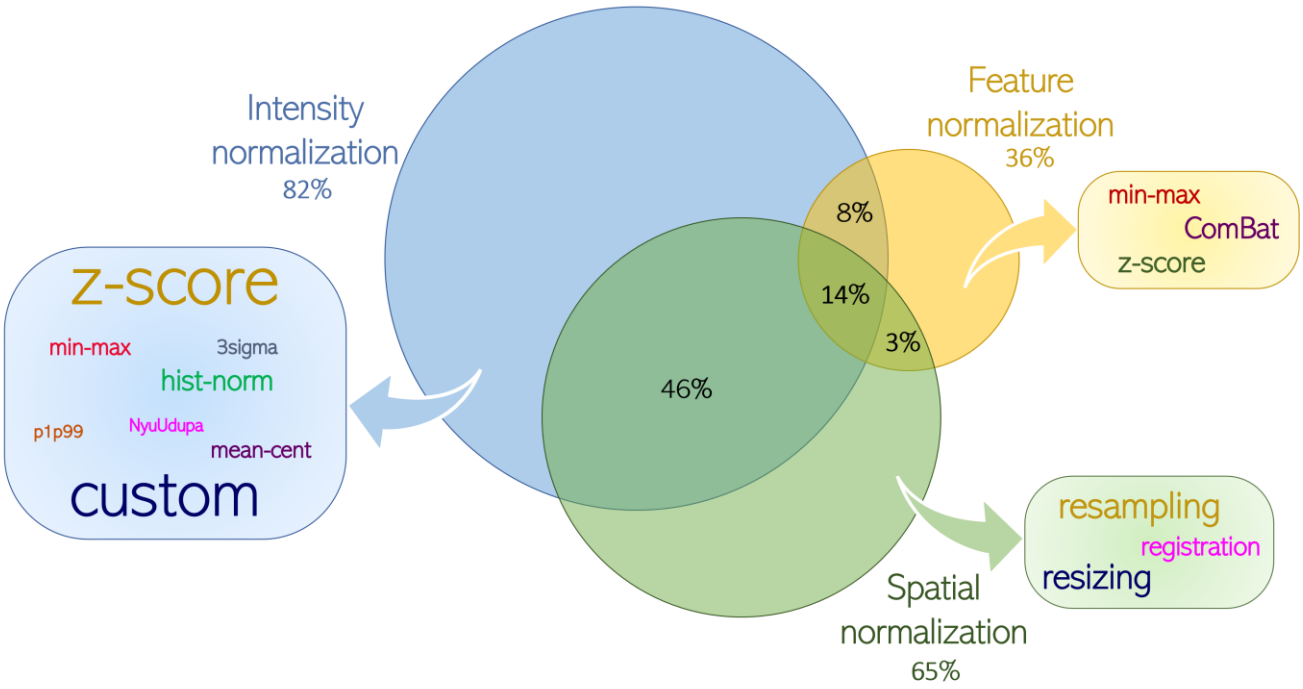

Fig. 4: This is the knowledge map of the most used normalizations among the included papers, showing the main approaches and methods followed. Each node represents a normalization approaches, while the boxes include the methods adopted. For each node it is showed the rate of each approach separately, and the rate of multi-approach normalizations. All rates are evaluated with respect to the amount of the included papers. The sizes of the words denote the number of publications related to the method. The dark blue line groups the total amount of the included papers.

normalized. Therefore, it requires available and labelled data from all centers/scanners included in the dataset, not allowing the direct application of previously determined normalization parameters on external validation cohorts [8]. Of note, 5/27 papers declared they applied a features normalization approach without specifying which method, and 1/27 combined both *ComBat* and *z-score*. Observing the S-Tables I-V, the *z-score* is mostly applied on the rectum (3/6), the *min-max* on the prostate (3/5), while the *ComBat* on the female pelvis (4/7).

D. Meta-Analysis

The 22 papers included in the meta-analysis are reported in the supplementary tables in bold. We did not perform the meta-analysis on detection, since there were not enough studies that could be included for this purpose.

Concerning the 16 papers addressing the characterization (fig. 5), the studies are statistically heterogeneous (heterogeneity=85%, *p-value*<0.01), and differences are partially explainable using the organs as the moderator (residual heterogeneity: 51%, *p-value*=0.02).

Considering each organ separately, only the kidney showed high significant heterogeneity between the two included studies (*p-value*<0.01), which differ in that one of them applies features normalization in addition to spatial and intensity normalizations. This behavior might suggest that in this case normalizing the features does not improve the performance (AUC=0.60 vs AUC=0.87), however, the sample size is very small to strongly assume this point. It is noteworthy that the papers related to the liver, rectal and prostate show very low heterogeneity (I^2 = 0% for the three organs) even though they applied very different normalization approaches. This might suggest that there is no preferable normalization strategy when dealing with these organs. Considering average results, two organs reach an average AUC higher than the overall pooled AUC, i.e., female pelvis (0.88 vs 0.81) and prostate (0.82 vs 0.81). It is noteworthy that, in studies regarding the female pelvis papers, *ComBat* feature normalization seems to slightly increase performances (AUC=0.86 and 0.91 vs 0.85). However, there is not a clear correlation between the used approach and performance, therefore it is not possible to define a pathway for MRI variability reduction.

Concerning the 6 papers addressing the prediction of response to therapy (fig. 6), the overall study heterogeneity is 98% (*p-value*<0.01), partially explainable using the organs as the moderator (residual heterogeneity: 83%). In this case, the performances of the two groups were statistically different (*p-value*<0.01), and in particular results on female pelvis were higher than rectal cancer (AUC = 0.96 vs 0.72, respectively), but both groups showed a heterogeneity higher than 75%. Focusing on the rectal area, all papers applied both spatial and intensity methods, except Song *et al*. [69] which performed also features normalization, yielding more robust performances on a larger population. Regarding the female pelvis, we observed that the two included studies applied intensity normalization using the *z-score* method, in combination with either spatial or feature normalization. In particular, the latter approach allowed reaching higher results (AUC=0.99 vs AUC=0.92). However, since there were only two papers in this group, we cannot strongly assume that the combination of *z-score* and features normalization is more suitable for this organ.

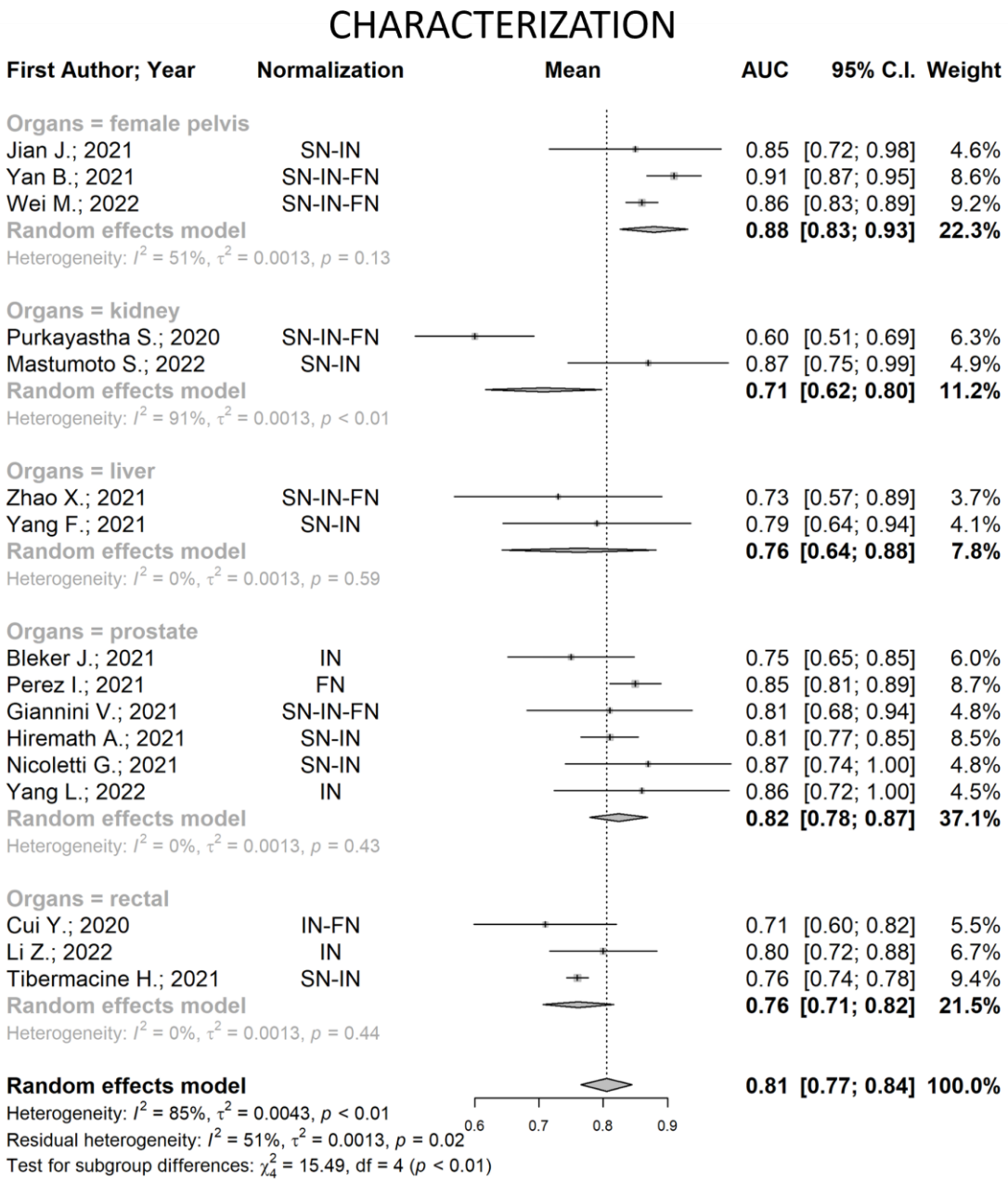

Fig. 5: Forest plot of the studies addressing characterization aim for the pooled area under the curve (AUC) and 95% confidence interval (CI). Horizontal CI lines represent 95% confidence interval of the point estimates, while the vertical dash represents the overall pooled AUC. The diamond represents the pooled AUC and its 95% CI obtained for each subgroup and considering the 14 papers all together.
Note: Intensity Normalization (IN), Spatial Normalization (SN), Feature Normalization (FN).

Considering all the above results, it is not possible to extract clear indications on the type of pre-processing for reducing abdominal MRI variability, mainly because we found a multiplicity of algorithms and pipelines applied. Moreover, no definite correlation between normalization and performance emerges from the meta-analyses, since most of the considered subgroups were very small. Both aspects could be partially associated to the quite recent interest in multi-center studies concerning abdominal MRI, increased only from 2020.

The prostate was the only organ showing a low heterogeneity with a sufficiently high number of included papers: in this case, the most used methods for the characterization aim are *resampling* for spatial and *min-max* for features normalizations. Conversely, intensity normalization was preferably applied using *custom* algorithms, tailored by each research group considering the characteristics of the organ, i.e., its heterogeneity and the presence of small lesion difficult to be detected even by experts, and clinical task. For these reasons, we strongly recommend to carefully analyze the available images and aim.

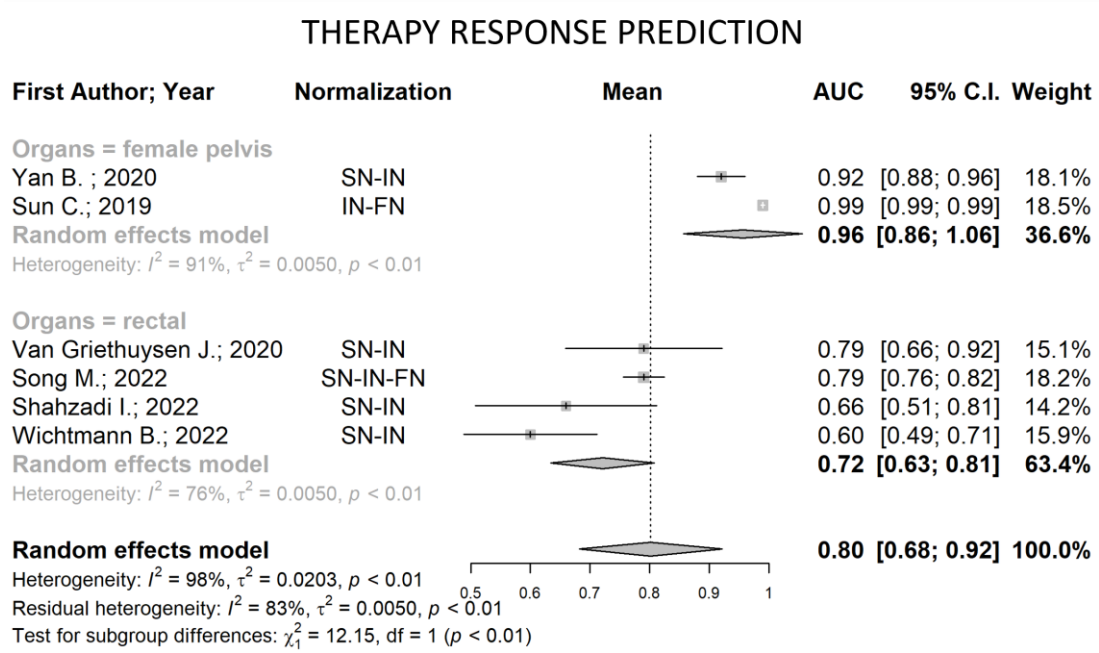

Fig. 6: Forest plot of the studies addressing the therapy response prediction aim for the pooled area under the curve (AUC) and 95% CI. Horizontal lines represent 95% confidence interval of the point estimates, while the vertical dash represents AUC of individual studies. The diamond means the pooled AUC of each subgroup and of all 6 papers.
Note: Intensity Normalization (IN), Spatial Normalization (SN), Feature Normalization (FN).

Regarding both the rectal are and female pelvis, we noticed that *resampling and resize* have become a standard step since 2021. Moreover, the combination with *z-score* intensity normalization seems to lead to higher results for characterization and therapy response aims.

In recent years, several efforts have been made toward the implementation of new approaches based on DL algorithms, i.e., transfer learning, data synthesis, thanks to their capability to learn from a given datasets [51], [82], [83]. Despite recent success, DL is not always a feasible approach for every clinical aim, since it requires a large amount of well annotated data. This is not a straightforward task for different reasons, including the lack of data for rare diseases, and lack of clinical check on the reference standard of most publicly available datasets.

This systematic review highlighted the shortage of evidences and guidelines on normalization approaches for abdominal MRI, differently from the agreed pipelines related to the brain [14], [97]. Up to now, an initial agreement was obtained by introducing Image Biomarker Standardization Initiative (IBSI), a protocol which works towards standardizing the extraction of radiomics features but does not suggest any pre-processing pipelines [98].

Several insights emerge from our analyses. First, it could be useful to evaluate how the different normalizations really affect the model performances. In this review, it was not possible to carry out this evaluation per each study since most of them did not present the results without applying the normalization. Then, despite being able to identify the most commonly used normalization approaches, we could not provide precise explanations of the reason behind, since the majority of the collected papers did not justify the choice or provided evidences of pros and cons of different methodologies. Finally, the number of studies included in the two meta-analyses is lower than the reviewed articles, since not all of them evaluated the performances using the same metric and externally validated their models.

Our preliminary findings should be further validated using a larger amount of paper. This could be achieved by including other normalization methods applied on different AI-based system development steps, such as feature selection and dimensionality reduction, [99], and feature filtering [100].

IV. Conclusion

Recently the need to define a suitable and useful normalization method for the reduction of multi-center MRI database variability for all abdominal organs has increased. Thanks to the findings obtained by the systematic review and meta-analyses carried out on subgroups of papers, we observed that there are some commonly used approaches, but not clear guidelines on different methodologies. Therefore, the definition of an abdominal pre-processing pipeline is still ongoing research, and it is of crucial importance to keep working on defining a proper methodology to reduce the multi-center database variability. In conclusion, we suggest carefully selecting the proper normalization approach considering the MRI database provided, the clinical aim, and the radiomics model.

Supplementary Materials

In the supplementary materials the reader will find five comparison tables collecting all information obtained by the paper analysis. Each table refers to different abdominal organs.

REFERENCES

[1] P. Lambin *et al.*, "Radiomics: The bridge between medical imaging and personalized medicine," *Nat. Rev. Clin. Oncol.*, vol. 14, no. 12, pp. 749–762, 2017.

[2] P. Lambin *et al.*, "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer*, 2012.

[3] V. Giannini *et al.*, "A Fully Automatic Artificial Intelligence System Able to Detect and Characterize Prostate Cancer Using Multiparametric MRI: Multicenter and Multi-Scanner Validation," *Front. Oncol.*, vol. 11, no. October, pp. 1–13, 2021.

[4] T. Upadhaya *et al.*, "Comparison of radiomics models built through machine learning in a multicentric context with independent testing: Identical data, similar algorithms, different methodologies," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 192–200, 2019.

[5] L. Shao *et al.*, "Multiparametric MRI and Whole Slide Image-Based Pretreatment Prediction of Pathological Response to Neoadjuvant Chemoradiotherapy in Rectal Cancer: A Multicenter Radiopathomic Study," *Ann. Surg. Oncol.*, vol. 27, no. 11, pp. 4296–4306, 2020.

[6] X. Zhao *et al.*, "Radiomics Based on Contrast-Enhanced MRI in Differentiation Between Fat-Poor Angiomyolipoma and Hepatocellular Carcinoma in Noncirrhotic Liver: A Multicenter Analysis," *Front. Oncol.*, vol. 11, no. October, pp. 1–11, 2021.

[7] A. Defeudis *et al.*, "Standardization of CT radiomics features for multi-center analysis: Impact of software settings and parameters," *Phys. Med. Biol.*, vol. 65, no. 19, 2020.

[8] R. Da-ano *et al.*, "Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020.

[9] R. Da-Ano, D. Visvikis, and M. Hatt, "Harmonization strategies for multicenter radiomics investigations," *Phys. Med. Biol.*, vol. 65, no. 24, 2020.

[10] E. Scalco and G. Rizzo, "Texture analysis of medical images for radiotherapy applications," *Br. J. Radiol.*, vol. 90, no. 1070, 2017.

[11] R. Aljundi, J. Lehaire, F. Prost-Boucle, O. Rouvière, and C. Lartizien, "Transfer learning for prostate cancer mapping based on multicentric MR imaging databases," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9487, pp. 74–82, 2015.

[12] F. Orlhac *et al.*, "A guide to ComBat harmonization of imaging biomarkers in multicenter studies," *J. Nucl. Med.*, vol. 63, no. 2, 2022.

[13] J. S. Song, H. S. Kwak, J. H. Byon, and G. Y. Jin, "Diffusion-weighted MR imaging of upper abdominal organs at different time points: Apparent diffusion coefficient normalization using a reference organ," *J. Magn. Reson. Imaging*, vol. 45, no. 5, pp. 1494–1501, 2017.

[14] B. Y. Park, K. Byeon, and H. Park, "FuNP (fusion of neuroimaging preprocessing) pipelines: A fully automated preprocessing software for functional magnetic resonance imaging," *Front. Neuroinform.*, vol. 13, no. February, pp. 1–14, 2019.

[15] N. Wijethilake, D. Meedeniya, C. Chitraranjan, M. Islam, and H. Ren, "Glioma survival analysis empowered with data engineering-a survey," *IEEE Access*, vol. 9, pp. 43168–43191, 2021.

[16] C. L. Ramspek, K. J. Jager, F. W. Dekker, C. Zoccali, and M. Van DIepen, "External validation of prognostic models: What, why, how, when and where?," *Clin. Kidney J.*, vol. 14, no. 1, pp. 49–58, 2021.

[17] C. H. Lee, S. Cook, J. S. Lee, and B. Han, "Comparison of Two Meta-Analysis Methods: Inverse-Variance-Weighted Average and Weighted Sum of Z-Scores," *Genomics Inform.*, vol. 14, no. 4, p. 173, 2016.

[18] W. Viechtbauer, "Conducting meta-analyses in R with the metafor," *J. Stat. Softw.*, vol. 36, no. 3, pp. 1–48, 2010.

[19] G. A. Nketiah *et al.*, "Utility of T2-weighted MRI texture analysis in assessment of peripheral zone prostate cancer aggressiveness: a single-arm, multicenter study," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, 2021.

[20] I. Montoya Perez *et al.*, "Detection of Prostate Cancer Using Biparametric Prostate MRI, Radiomics, and Kallikreins: A Retrospective Multicenter Study of Men With a Clinical Suspicion of Prostate Cancer," *J. Magn. Reson. Imaging*, vol. 55, no. 2, pp. 465–477, 2022.

[21] S. E. Viswanath *et al.*, "Comparing radiomic classifiers and classifier ensembles for detection of peripheral zone prostate tumors on T2-weighted MRI: A multi-site study," *BMC Med. Imaging*, vol. 19, no. 1, pp. 1–12, 2019.

[22] A. Hiremath *et al.*, "An integrated nomogram combining deep learning, Prostate Imaging–Reporting and Data System (PI-RADS) scoring, and clinical variables for identification of clinically significant prostate cancer on biparametric MRI: a retrospective multicentre study," *Lancet Digit. Heal.*, vol. 3, no. 7, pp. e445–e454, 2021.

[23] Q. Liu, Q. Dou, L. Yu, and P. A. Heng, "MS-Net: Multi-Site Network for Improving Prostate Segmentation with Heterogeneous MRI Data," *IEEE Trans. Med. Imaging*, vol. 39, no. 9, pp. 2713–2724, 2020.

[24] G. Nicoletti *et al.*, "Virtual biopsy in prostate cancer: Can machine learning distinguish low and high aggressive tumors on MRI," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 3374–3377, 2021.

[25] H. Bagher-Ebadian *et al.*, "Detection of Dominant Intra-prostatic Lesions in Patients With Prostate Cancer Using an Artificial Neural Network and MR Multi-modal Radiomics Analysis," *Front. Oncol.*, vol. 9, no. November, pp. 1–14, 2019.

[26] J. Bleker *et al.*, "A deep learning masked segmentation alternative to manual segmentation in biparametric MRI prostate cancer radiomics," *Eur. Radiol.*, vol. 32, no. 9, pp. 6526–6535, 2022.

[27] L. Hu, D. W. Zhou, X. Y. Guo, W. H. Xu, L. M. Wei, and J. G. Zhao, "Adversarial training for prostate cancer classification using magnetic resonance imaging," *Quant. Imaging Med. Surg.*, vol. 12, no. 6, pp. 3276–3287, 2022.

[28] Y. Yan *et al.*, "Deep learning with quantitative features of magnetic resonance images to predict biochemical recurrence of radical prostatectomy: A multi-center study," *Cancers (Basel).*, vol. 13, no. 12, 2021.

[29] A. Fernandez-Quilez, S. V. Larsen, M. Goodwin, T. O. Gulsrud, S. R. Kjosavik, and K. Oppedal, "Improving prostate whole gland segmentation in T2-weighted mri with synthetically generated data," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2021-April, pp. 1915–1919, 2021.

[30] N. Karani, E. Erdil, K. Chaitanya, and E. Konukoglu, "Test-time adaptable neural networks for robust medical image segmentation," *Med. Image Anal.*, vol. 68, p. 101907, 2021.

[31] L. Shao *et al.*, "Patient-level grading prediction of prostate cancer from mp-MRI via GMINet," *Comput. Biol. Med.*, vol. 150, no. September, p. 106168, 2022.

[32] G. Jing *et al.*, "Prediction of clinically significant prostate cancer with a multimodal MRI-based radiomics nomogram," *Front. Oncol.*, vol. 12, no. July, pp. 1–11, 2022.

[33] L. Yang *et al.*, "Radiomic Machine Learning and External Validation Based on 3.0 T mpMRI for Prediction of Intraductal Carcinoma of Prostate With Different Proportion," *Front. Oncol.*, vol. 12, no. June, pp. 1–9, 2022.

[34] J. Bleker *et al.*, "Single-center versus multi-center biparametric MRI radiomics approach for clinically significant peripheral zone prostate cancer," *Insights Imaging*, vol. 12, no. 1, 2021.

[35] D. J. Winkel *et al.*, "Autonomous detection and classification of pi-rads lesions in an mri screening population incorporating multicenter-labeled deep learning and biparametric imaging: Proof of concept," *Diagnostics*, vol. 10, no. 11, 2020.

[36] C. V. Dinh *et al.*, "Multi-center validation of prostate tumor localization using multi-parametric MRI and prior knowledge."

[37] L. Shao *et al.*, "Radiologist-like artificial intelligence for grade group prediction of radical prostatectomy for reducing upgrading and downgrading from biopsy," *Theranostics*, vol. 10, no. 22, pp. 10200–10212, 2020.

[38] R. Cuocolo *et al.*, "MRI index lesion radiomics and machine learning for detection of extraprostatic extension of disease: a multicenter study," *Eur. Radiol.*, vol. 31, no. 10, pp. 7575–7583, 2021.

[39] J. M. Castillo T. *et al.*, "A multi-center, multi-vendor study to evaluate the generalizability of a radiomics model for classifying prostate cancer: High grade vs. low grade," *Diagnostics*, vol. 11, no. 2, 2021.

[40] S. B. Ginsburg *et al.*, "Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: Preliminary findings from a multi-institutional study," *J. Magn. Reson. Imaging*, vol. 46, no. 1, pp. 184–193, 2017.

[41] J. Jian *et al.*, "MR image-based radiomics to differentiate type I and type II epithelial ovarian cancers," *Eur. Radiol.*, vol. 31, no. 1, pp. 403–410, 2021.

[42] K. Nagawa *et al.*, "Diagnostic utility of a conventional MRI-based analysis and texture analysis for discriminating between ovarian thecoma-fibroma groups and ovarian granulosa cell tumors," *J. Ovarian Res.*, vol. 15, no. 1, pp. 1–14, 2022.

[43] J. Shi *et al.*, "MRI-based intratumoral and peritumoral radiomics on prediction of lymph-vascular space invasion in cervical cancer: A multi-center study," *Biomed. Signal Process. Control*, vol. 72, no. PB, p. 103373, 2022.

[44] Q. Bi *et al.*, "Different multiparametric MRI-based radiomics models for differentiating stage IA endometrial cancer from benign endometrial lesions: A multicenter study," *Front. Oncol.*, vol. 12, no. August, pp. 1–13, 2022.

[45] B. C. Yan *et al.*, "Preoperative Assessment for High-Risk Endometrial Cancer by Developing an MRI- and Clinical-Based Radiomics Nomogram: A Multicenter Study," *J. Magn. Reson. Imaging*, vol. 52, no. 6, pp. 1872–1882, 2020.

[46] C. Sun *et al.*, "Radiomic analysis for pretreatment prediction of response to neoadjuvant chemotherapy in locally advanced cervical cancer: A multicentre study," *EBioMedicine*, vol. 46, pp. 160–169, 2019.

[47] F. Lucia *et al.*, "External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 46, no. 4, pp. 864–877, 2019.

[48] R. Da-Ano *et al.*, "A transfer learning approach to facilitate ComBat-based harmonization of multicentre radiomic features in new datasets," *PLoS One*, vol. 16, no. 7 July, pp. 1–19, 2021.

[49] B. C. Yan *et al.*, "Radiologists with MRI-based radiomics aids to predict the pelvic lymph node metastasis in endometrial cancer: a multicenter study," *Eur. Radiol.*, vol. 31, no. 1, pp. 411–422, 2021.

[50] Y. Li *et al.*, "MRI-Based Machine Learning for Differentiating Borderline From Malignant Epithelial Ovarian Tumors: A Multicenter Study," *J. Magn. Reson. Imaging*, vol. 52, no. 3, pp. 897–904, 2020.

[51] X. Chang *et al.*, "Self-supervised learning for multi-center magnetic resonance imaging harmonization without traveling phantoms," *Phys. Med. Biol.*, vol. 67, no. 14, 2022.

[52] M. Wei *et al.*, "T2-weighted MRI-based radiomics for discriminating between benign and borderline epithelial ovarian tumors: a multicenter study," *Insights Imaging*, vol. 13, no. 1, 2022.

[53] D. Cusumano *et al.*, "A field strength independent MR radiomics model to predict pathological complete response in locally advanced rectal cancer," *Radiol. Medica*, vol. 126, no. 3, pp. 421–429, 2021.

[54] I. Shahzadi *et al.*, "Analysis of MRI and CT-based radiomics features for personalized treatment in locally advanced rectal cancer and external validation of published radiomics models," *Sci. Rep.*, vol. 12, no. 1, pp. 1–15, 2022.

[55] Z. Li *et al.*, "The feasibility of MRI-based radiomics model in presurgical evaluation of tumor budding in locally advanced rectal cancer," *Abdom. Radiol.*, vol. 47, no. 1, pp. 56–65, 2022.

[56] B. D. Wichtmann *et al.*, "Are We There Yet? The Value of Deep Learning in a Multicenter Setting for Response Prediction of Locally Advanced Rectal Cancer to Neoadjuvant Chemoradiotherapy," *Diagnostics*, vol. 12, no. 7, 2022.

[57] F. Knuth *et al.*, "MRI-based automatic segmentation of rectal cancer using 2D U-Net on two independent cohorts," *Acta Oncol. (Madr).*, vol. 0, no. 0, pp. 1–9, 2021.

[58] J. Panic *et al.*, "A fully automatic deep learning algorithm to segment rectal Cancer on MR images: a multi-center study," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2022-July, pp. 5066–5069, 2022.

[59] A. Defeudis *et al.*, "MRI-based radiomics to predict response in locally advanced rectal cancer: comparison of manual and automatic segmentation on external validation in a multicentre study," *Eur. Radiol. Exp.*, vol. 6, no. 1, 2022.

[60] Y. Xiang *et al.*, "MRI-based radiomics to predict neoadjuvant chemoradiotherapy outcomes in locally advanced rectal cancer: A multicenter study," *Clin. Transl. Radiat. Oncol.*, vol. 38, no. November 2022, pp. 175–182, 2023.

[61] H. Tibermacine *et al.*, "Radiomics modelling in rectal cancer to predict disease-free survival: evaluation of different approaches," *Br. J. Surg.*, vol. 108, no. 10, pp. 1243–1250, 2021.

[62] Z. Liu *et al.*, "Predicting distant metastasis and chemotherapy benefit in locally advanced rectal cancer," *Nat. Commun.*, vol. 11, no. 1, pp. 1–11, 2020.

[63] H. Shaish *et al.*, "Radiomics of MRI for pretreatment prediction of pathologic complete response, tumor regression grade, and neoadjuvant rectal score in patients with locally advanced rectal cancer undergoing neoadjuvant chemoradiation: an international multicenter study," *Eur. Radiol.*, vol. 30, no. 11, pp. 6263–6273, 2020.

[64] J. J. M. van Griethuysen *et al.*, "Radiomics performs comparable to morphologic assessment by expert radiologists for prediction of response to neoadjuvant chemoradiotherapy on baseline staging MRI in rectal cancer," *Abdom. Radiol.*, vol. 45, no. 3, pp. 632–643, 2020.

[65] X. Liu *et al.*, "Deep learning radiomics-based prediction of distant metastasis in patients with locally advanced rectal cancer after neoadjuvant chemoradiotherapy: A multicentre study," *EBioMedicine*, vol. 69, p. 103442, 2021.

[66] N. Giraud *et al.*, "MRI-based radiomics input for prediction of 2-year disease recurrence in anal squamous cell carcinoma," *Cancers (Basel).*, vol. 13, no. 2, pp. 1–11, 2021.

[67] H. T. Zhu, X. Y. Zhang, Y. J. Shi, X. T. Li, and Y. S. Sun, "A Deep Learning Model to Predict the Response to Neoadjuvant Chemoradiotherapy by the Pretreatment Apparent Diffusion Coefficient Images of Locally Advanced Rectal Cancer," *Front. Oncol.*, vol. 10, no. October, pp. 1–8, 2020.

[68] Y. Cui *et al.*, "Development and validation of a MRI-based radiomics signature for prediction of KRAS mutation in rectal cancer," *Eur. Radiol.*, vol. 30, no. 4, pp. 1948–1958, 2020.

[69] M. Song *et al.*, "MRI radiomics independent of clinical baseline characteristics and neoadjuvant treatment modalities predicts response to neoadjuvant therapy in rectal cancer," *Br. J. Cancer*, vol. 127, no. 2, pp. 249–257, 2022.

[70] X. Zhang *et al.*, "Effective staging of fibrosis by the selected texture features of liver: Which one is better, CT or MR imaging?," *Comput. Med. Imaging Graph.*, vol. 46, no. September, pp. 227–236, 2015.

[71] F. Yang *et al.*, "MRI-Radiomics Prediction for Cytokeratin 19-Positive Hepatocellular Carcinoma: A Multicenter Study," *Front. Oncol.*, vol. 11, no. August, pp. 1–8, 2021.

[72] Y. Kuang *et al.*, "MRI-Based Radiomics: Nomograms predicting the short-term response after transcatheter arterial chemoembolization (TACE) in hepatocellular carcinoma patients with diameter less than 5 cm," *Abdom. Radiol.*, vol. 46, no. 8, pp. 3772–3789, 2021.

[73] Y. Cho *et al.*, "Computer-aided hepatocellular carcinoma detection on the hepatobiliary phase of gadoxetic acid-enhanced magnetic resonance imaging using a convolutional neural network: Feasibility evaluation with multi-sequence data," *Comput. Methods Programs Biomed.*, vol. 225, p. 107032, 2022.

[74] J. Kim, J. H. Min, S. K. Kim, S. Y. Shin, and M. W. Lee, "Detection of Hepatocellular Carcinoma in Contrast-Enhanced Magnetic Resonance Imaging Using Deep Learning Classifier: A Multi-Center Retrospective Study," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, 2020.

[75] Y. Di Chen *et al.*, "Radiomics and nomogram of magnetic resonance imaging for preoperative prediction of microvascular invasion in small hepatocellular carcinoma," *World J. Gastroenterol.*, vol. 28, no. 31, pp. 4399–4416, 2022.

[76] I. L. Xi *et al.*, "Deep learning to distinguish benign from malignant renal lesions based on routine MR imaging," *Clin. Cancer Res.*, vol. 26, no. 8, pp. 1944–1952, 2020.

[77] S. Purkayastha *et al.*, "Differentiation of low and high grade renal cell carcinoma on routine MRI with an

externally validated automatic machine learning algorithm," *Sci. Rep.*, vol. 10, no. 1, pp. 1–8, 2020.

[78] Y. Zhao *et al.*, "Deep Learning Based on MRI for Differentiation of Low- and High-Grade in Low-Stage Renal Cell Carcinoma," *J. Magn. Reson. Imaging*, vol. 52, no. 5, pp. 1542–1549, 2020.

[79] J. W. Choi *et al.*, "Preoperative prediction of the stage, size, grade, and necrosis score in clear cell renal cell carcinoma using MRI-based radiomics," *Abdom. Radiol.*, vol. 46, no. 6, pp. 2656–2664, 2021.

[80] S. Matsumoto *et al.*, "Utility of radiomics features of diffusion-weighted magnetic resonance imaging for differentiation of fat-poor angiomyolipoma from clear cell renal cell carcinoma: model development and external validation," *Abdom. Radiol.*, vol. 47, no. 6, pp. 2178–2186, 2022.

[81] H. Wang *et al.*, "Elaboration of a multisequence MRI-based radiomics signature for the preoperative prediction of the muscle-invasive status of bladder cancer: a double-center study," *Eur. Radiol.*, vol. 30, no. 9, pp. 4816–4827, 2020.

[82] Y. Zou *et al.*, "Multi-task deep learning based on T2-Weighted Images for predicting Muscular-Invasive Bladder Cancer," *Comput. Biol. Med.*, vol. 151, no. PA, p. 106219, 2022.

[83] X. Gao and X. Wang, "Performance of deep learning for differentiating pancreatic diseases on contrast-enhanced magnetic resonance imaging: A preliminary study," *Diagn. Interv. Imaging*, vol. 101, no. 2, pp. 91–100, 2020.

[84] S. Cui *et al.*, "Radiomic nomogram based on MRI to predict grade of branching type intraductal papillary mucinous neoplasms of the pancreas: a multicenter study," *Cancer Imaging*, vol. 21, no. 1, pp. 1–13, 2021.

[85] A. Crombé *et al.*, "Progressive Desmoid Tumor : Study by the French Sarcoma Group," vol. 1, no. December, pp. 1–10, 2020.

[86] A. Crombé *et al.*, "Intensity harmonization techniques influence radiomics features and radiomics-based predictions in sarcoma patients," *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, 2020.

[87] Q. Wu *et al.*, "Radiomics analysis of placenta on T2WI facilitates prediction of postpartum haemorrhage: A multicentre study," *EBioMedicine*, vol. 50, pp. 355–365, 2019.

[88] Z. Ye, R. Xuan, M. Ouyang, Y. Wang, J. Xu, and W. Jin, "Prediction of placenta accreta spectrum by combining deep learning and radiomics using T2WI: a multicenter study," *Abdom. Radiol.*, vol. 47, no. 12, pp. 4205–4218, 2022.

[89] S. Liu *et al.*, "Deep learning radiomic nomogram to predict recurrence in soft tissue sarcoma: a multi-institutional study," *Eur. Radiol.*, vol. 32, no. 2, pp. 793–805, 2022.

[90] A. E. Kavur *et al.*, "CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation," *Med. Image Anal.*, vol. 69, 2021.

[91] T. Hulsen, "An overview of publicly available patient-centered prostate cancer datasets," *Transl. Androl. Urol.*, vol. 8, no. S1, pp. S64–S77, 2019.

[92] G. Lemaître, R. Martí, J. Freixenet, J. C. Vilanova, P. M. Walker, and F. Meriaudeau, "Computer-Aided Detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review," *Comput. Biol. Med.*, vol. 60, pp. 8–31, 2015.

[93] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, and E. I. Chang, "DEEP LEARNING OF FEATURE REPRESENTATION WITH MULTIPLE INSTANCE LEARNING FOR MEDICAL IMAGE ANALYSIS," *IEEE Int. Conf. Acoust. Speech Signal Process.*, no. 1, pp. 1645–1649, 2014.

[94] D. Xiao *et al.*, "Diagnosis of Invasive Meningioma Based on Brain-Tumor Interface Radiomics Features on Brain MR Images: A Multicenter Study," *Front. Oncol.*, vol. 11, no. August, pp. 1–13, 2021.

[95] X. Sun *et al.*, "Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions," *Biomed. Eng. Online*, vol. 14, no. 1, pp. 1–17, 2015.

[96] L. G. Nyú and J. K. Udupa, "On standardizing the MR image intensity scale," *Magn. Reson. Med.*, vol. 42, no. 6, pp. 1072–1081, 1999.

[97] A. Carré *et al.*, "Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics," *Sci. Rep.*, vol. 10, no. 1, pp. 1–15, 2020.

[98] L. S. Zwanenburg A, Leger S, Vallières M, "Image biomarke standardization initiative," *arXiv Prepr. arXiv1612.07003*, 2016.

[99] M. Götz and K. H. Maier-Hein, "Optimal Statistical Incorporation of Independent Feature Stability Information into Radiomics Studies," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, 2020.

[100] Haueise T, Liebgott A, Yang B. A Comparative Study on the Potential of Unsupervised Deep Learning-based Feature Selection in Radiomics. Annu Int Conf IEEE Eng Med Biol Soc. 2022 Jul;2022:541-544.